

An inertial primal-dual fixed point algorithm for composite optimization problems

Meng Wen^{1,2}, Yu-Chao Tang³, Jigen Peng^{1,2}

1. School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, P.R. China
2. Beijing Center for Mathematics and Information Interdisciplinary Sciences, Beijing, P.R. China
3. Department of Mathematics, NanChang University, Nanchang 330031, P.R. China

Abstract In this paper, we consider an inertial primal-dual fixed point algorithm (IPDFP) to compute the minimizations of the following Problem (1.1). This is a full splitting approach, in the sense that the nonsmooth functions are processed individually via their proximity operators. The convergence of the IPDFP is obtained by reformulating the Problem (1.1) to the sum of three convex functions. This work brings together and notably extends several classical splitting schemes, like the primal-dual method proposed by Chambolle and Pock, and the recent proximity algorithms of Charles A. et al designed for the L_1 /TV image denoising model. The iterative algorithm is used for solving nondifferentiable convex optimization problems arising in image processing. The experimental results indicate that the proposed IPDFP iterative algorithm performs well with respect to state-of-the-art methods.

Keywords: composite optimization; operator splitting; proximity operator; inertial
MR(2000) Subject Classification 47H09, 90C25,

1 Introduction

The purpose of this paper is to designing and discussing an efficient algorithmic framework with inertial version for minimizing the following problem

$$\min_{x \in \mathcal{X}} \sum_{i=1}^m F_i(K_i x) + G(x), \quad (1.1)$$

* Corresponding author.

E-mail address: wen5495688@163.com

where \mathcal{X} and $\{\mathcal{Y}_i\}_{i=1}^m$ are Hilbert spaces, and $G \in \Gamma_0(\mathcal{X})$, $F_i \in \Gamma_0(\mathcal{Y}_i)$ respectively; $K_i : \mathcal{X} \rightarrow \mathcal{Y}_i$ be a continuous linear operator, for $i = 1, \dots, m$. Here and in what follows, for a real Hilbert space H , $\Gamma_0(H)$ denotes the collection of all proper lower semi-continuous convex functions from H to $(-\infty, +\infty]$.

To that end let us first rephrase Problem (1.1) as

$$\min_{x \in \mathcal{X}^m} \sum_{i=1}^m (F_i(K_i x_i) + \frac{1}{m} G(x_i)) + \delta_C(x), \quad (1.2)$$

where the Hilbert space $\mathcal{X}^m := \mathcal{X}_1 \times \dots \times \mathcal{X}_m$, equipped with the inner product $\langle x, x' \rangle := \sum_{i=1}^m \langle x_i, x'_i \rangle$, and $K_i : \mathcal{X}_i \rightarrow \mathcal{Y}_i$ be a continuous linear operator. The notation x_i represents the i -th component of any $x \in \mathcal{X}^m$, and C is the space of vectors $x \in \mathcal{X}^m$ such that $x_1 = \dots = x_m$. We define the linear function $K : \mathcal{X}^m \rightarrow \mathcal{Y}^m$ by $Kx := (K_1 x_1, \dots, K_m x_m)$, and we set $F(Kx) = \sum_{i=1}^m (F_i(K_i x_i))$ and $\bar{G}(x) = \sum_{i=1}^m \frac{1}{m} G(x_i)$. Then the Problem (1.2) can be rewritten as

$$\min_{x \in \mathcal{X}^m} F(Kx) + \bar{G}(x) + \delta_C(x), \quad (1.3)$$

In order to solve the above problem, first we consider the following general optimization problem:

$$\min_{x \in \mathcal{X}} F(Kx) + \bar{G}(x) + H(x), \quad (1.4)$$

where $F \in \Gamma_0(\mathcal{Y}^m)$, $\bar{G}, H \in \Gamma_0(\mathcal{X}^m)$.

When \bar{G} is differentiable on \mathcal{X}^m and its gradient $\nabla \bar{G}$ is β -Lipschitz continuous, for some $\beta \in [0, +\infty[$; that is,

$$\|\nabla \bar{G}(x) - \nabla \bar{G}(x')\| \leq \beta \|x - x'\|, \forall (x, x') \in \mathcal{X}^m \times \mathcal{X}^m.$$

the primal-dual method in [1] can be used to solve (1.4). Elaborating on the method introduced by Laurent Condat in [1] and the method given by Nesterov in [5], we provide an iterative algorithm for solving (1.4) which we refer to as IPDFP (inertial primal-dual fixed point algorithm). we will give the details of our method in the section 3.

Despite the form of (1.1) is simply, many problems in image processing can be formulated it. For instance, the following $L_1/\varphi \circ B$ model. This model minimizes the

sum of the l_1 fidelity term and the composition of a convex function with a matrix and includes the L_1 /TV denoising model and the L_1 /TV inpainting model as special cases.

$$\min_{x \in R^2} \|x - b\|_1 + \delta_C(x) + \varphi(\circ B)(x), \quad (1.5)$$

where φ is a given convex on R^m and B is a given $m \times n$ matrix. For the anisotropic total-variation φ is the norm $\|\cdot\|_1$ while for the isotropic total-variation φ is a linear combination of the norm $\|\cdot\|_2$ in R^2 . The matrix B for the both cases is the first order difference matrix. For higher-order total-variations (see, e.g., [19-24]), B may be chosen to be a higher-order difference matrix. Problem (1.5) can be expressed in the form of (1.1) by setting $m = 2$, $G(x) = \|x - b\|_1$, $F_1 = \delta_C$, $K_1 = I$ and $F_2 = \varphi$, $K_2 = B$. One of the main difficulties in solving it is that F_i and G are non-differentiable. The case often occurs in many problems we are interested in.

In this paper, the contributions of us are the following aspects:

(I) We provide an inertial primal-dual fixed point algorithm to solve the general Problem (1.1), which is inspired by the primal-dual splitting method present by Laurent Condat [1] and the method introduced by Nesterov [5]. We refer to our algorithm as IPDFP. Firstly, when \bar{G} is differentiable and $\alpha_k = 0$ in our method, the primal-dual splitting method introduced by Laurent Condat [1] is a special case of our algorithm. Secondly, for $m = 1$ and $\alpha_k = 0$, it includes the well known first-order primal-dual algorithm proposed by Chambolle and Pock. Finally, when $m = 1$ and $K_1 = I$, we can obtain the inertial forward-backward algorithm introduced by Dirk A. Lorenz and Thomas Pock [18].

(II) Based on the idea of preconditioning techniques, we propose simple and easy to compute diagonal preconditioners for which convergence of the algorithm is guaranteed without the need to compute any step size parameters and it leaves the computational complexity of the iterations basically unchanged. As a by-product, we show that for a certain instance of the preconditioning, the proposed algorithm is equivalent to the old and widely unknown primal-dual algorithm.

(III) With the idea of the inertial version of the Krasnosel'skii-Mann iterations algorithm for approximating the set of fixed points of a nonexpansive operator and the particular inner product defined by a symmetric positive definite map P which can be interpreted as a preconditioner or variable metric, we prove the convergence of our

method.

The rest of this paper is organized as follows. In the next section, we introduce some notations used throughout in the paper. In section 3, we devote to introduce IPDFP and SIPDFP algorithm, and the relation between them, we also show how the IPDFP splits into SIPDFP and the convergence of proposed method. In section 4, we present the preconditioned primal-dual algorithm and give conditions under which convergence of the algorithm is guaranteed. We propose a family of simple and easy to compute diagonal preconditioners, which turn out to be very efficient on many problems. In the final section, we show the numerical performance and efficiency of propose algorithm through some examples in the context of large-scale l_1 -regularized logistic regression.

2 Preliminaries

Throughout the paper, we denote by $\langle \cdot, \cdot \rangle$ the inner product on \mathcal{X} and by $\|\cdot\|$ the norm on \mathcal{X} .

Assumption 2.1. *The infimum of Problem (1.4) is attained. Moreover, the following qualification condition holds*

$$0 \in ri(dom h - D dom g).$$

The dual problem corresponding to the primal Problem (1.4) is written

$$\min_{y \in \mathcal{Y}} (f + g)^*(-D^*y) + h^*(y),$$

where a^* denotes the Legendre-Fenchel transform of a function a and where D^* is the adjoint of D . With the Assumption 2.1, the classical Fenchel-Rockafellar duality theory [3], [11] shows that

$$\min_{x \in \mathbf{X}} f(x) + g(x) + (h \circ D)(x) = -\min_{y \in \mathcal{Y}} (f + g)^*(-D^*y) + h^*(y). \quad (2.1)$$

Definition 2.1. Let f be a real-valued convex function on \mathcal{X} , the operator prox_f is defined by

$$\text{prox}_f : \mathcal{X} \rightarrow \mathcal{X}$$

$$x \mapsto \arg \min_{y \in \mathcal{X}} f(y) + \frac{1}{2} \|x - y\|_2^2,$$

called the proximity operator of f .

Definition 2.2. Let A be a closed convex set of \mathcal{X} . Then the indicator function of A is defined as

$$\delta_A(x) = \begin{cases} 0, & \text{if } x \in A, \\ \infty, & \text{otherwise.} \end{cases}$$

It can easy see the proximity operator of the indicator function in a closed convex subset A can be reduced a projection operator onto this closed convex set A . That is,

$$\text{prox}_{\iota_A} = \text{proj}_A,$$

where proj is the projection operator of A .

Definition 2.3. (Nonexpansive operators and firmly nonexpansive operators [3]). Let \mathcal{H} be a Euclidean space (we refer to [3] for an extension to Hilbert spaces). An operator $T : \mathcal{H} \rightarrow \mathcal{H}$ is nonexpansive if and only if it satisfies

$$\|Tx - Ty\|_2 \leq \|x - y\|_2 \text{ for all } (x, y) \in \mathcal{H}^2.$$

T is firmly nonexpansive if and only if it satisfies one of the following equivalent conditions:

- (i) $\|Tx - Ty\|_2^2 \leq \langle Tx - Ty, x - y \rangle$ for all $(x, y) \in \mathcal{H}^2$;
- (ii) $\|Tx - Ty\|_2^2 = \|x - y\|_2^2 - \|(I - T)x - (I - T)y\|_2^2$ for all $(x, y) \in \mathcal{H}^2$.

It is easy to show from the above definitions that a firmly nonexpansive operator T is nonexpansive.

Definition 2.4. A mapping $T : \mathcal{H} \rightarrow \mathcal{H}$ is said to be an averaged mapping, if it can be written as the average of the identity I and a nonexpansive mapping; that is,

$$T = (1 - \alpha)I + \alpha S, \tag{2.2}$$

where α is a number in $]0, 1[$ and $S : \mathcal{H} \rightarrow \mathcal{H}$ is nonexpansive. More precisely, when (2.2) or the following inequality (2.2) holds, we say that T is α -averaged.

$$\|Tx - Ty\|^2 \leq \|x - y\|^2 - \frac{(1 - \alpha)}{\alpha} \|(I - T)x - (I - T)y\|^2, \forall x, y \in \mathcal{H}. \tag{2.3}$$

A 1-averaged operator is said non-expansive. A $\frac{1}{2}$ -averaged operator is said firmly non-expansive.

We refer the readers to [3] for more details. Let $M : \mathcal{H} \rightarrow \mathcal{H}$ be a set-valued operator. We denote by $\text{ran}(M) := \{v \in \mathcal{H} : \exists u \in \mathcal{H}, v \in Mu\}$ the range of M , by $\text{gra}(M) := \{(u, v) \in \mathcal{H}^2 : v \in Mu\}$ its graph, and by M^{-1} its inverse; that is, the set-valued operator with graph $(v, u) \in \mathcal{H}^2 : v \in Mu$. We define $\text{zer}(M) := \{u \in \mathcal{H} : 0 \in Mu\}$. M is said to be monotone if $\forall (u, u') \in \mathcal{H}^2, \forall (v, v') \in Mu \times Mu', \langle u - u', v - v' \rangle \geq 0$ and maximally monotone if there exists no monotone operator M' such that $\text{gra}(M) \subset \text{gra}(M') \neq \text{gra}(M)$.

The resolvent $(I + M)^{-1}$ of a maximally monotone operator $M : \mathcal{H} \rightarrow \mathcal{H}$ is defined and single-valued on \mathcal{H} and firmly nonexpansive. The subdifferential ∂J of $J \in \Gamma_0(\mathcal{H})$ is maximally monotone and $(I + \partial J)^{-1} = \text{prox}_J$.

Further, let us mention some classes of operators that are used in the paper. The operator A is said to be uniformly monotone if there exists an increasing function $\phi_A : [0; +1] \rightarrow [0; +1]$ that vanishes only at 0, and

$$\langle x - y, u - v \rangle \geq \phi_A(\|x - y\|), \forall (x, u), (y, v) \in \text{gra}(A). \quad (2.5)$$

Prominent representatives of the class of uniformly monotone operators are the strongly monotone operators. Let $\gamma > 0$ be arbitrary. We say that A is γ -strongly monotone, if $\langle x - y, u - v \rangle \geq \gamma\|x - y\|^2$, for all $(x, u), (y, v) \in \text{gra}(A)$.

Lemma 2.1. (see[2,9-10]). Let $(\varphi^k)_{k \in \mathbb{N}}$; $(\delta_k)_{k \in \mathbb{N}}$ and $(\alpha_k)_{k \in \mathbb{N}}$ be sequences in $[0; +1]$ such that $\varphi^{k+1} \leq \varphi^k + \alpha_k(\varphi^k - \varphi^{k-1}) + \delta_k$ for all $k \geq 1$, $\sum_{k \in \mathbb{N}} \delta_k < +\infty$ and there exists a real number α with $0 \leq \alpha_k \leq \alpha < 1$ for all $k \in \mathbb{N}$. Then the following hold:

- (i) $\sum_{k \geq 1} [\varphi^k - \varphi^{k-1}]_+ < +\infty$, where $[t]_+ = \max\{t, 0\}$;
- (ii) there exists $\varphi^* \in [0; +\infty)$ such that $\lim_{k \rightarrow +\infty} \varphi^k = \varphi^*$.

Lemma 2.2. ([4]). Let \tilde{M} be a nonempty closed and affine subset of a Hilbert space $\tilde{\mathcal{H}}$ and $T : \tilde{M} \rightarrow \tilde{M}$ a nonexpansive operator such that $\text{Fix}(T) \neq \emptyset$. Considering the following iterative scheme:

$$x^{k+1} = x^k + \alpha_k(x^k - x^{k-1}) + \rho_k[T(x^k + \alpha_k(x^k - x^{k-1})) - x^k - \alpha_k(x^k - x^{k-1})], \quad (2.6)$$

where $x^0; x^1$ are arbitrarily chosen in \tilde{M} , $(\alpha_k)_{k \in \mathbb{N}}$ is nondecreasing with $\alpha_1 = 0$ and $0 \leq \alpha_k \leq \alpha < 1$ for every $n \geq 1$ and $\rho, \theta, \hat{\delta} > 0$ are such that $\hat{\delta} > \frac{\alpha^2(1+\alpha)+\alpha\theta}{1-\alpha^2}$ and

$$0 < \rho \leq \rho_k < \frac{\hat{\delta} - \alpha[\alpha(1+\alpha) + \alpha\hat{\delta} + \theta]}{\hat{\delta}[1 + \alpha(1+\alpha) + \alpha\hat{\delta} + \theta]} \quad \forall k \geq 1.$$

Then the following statements are true:

- (i) $\sum_{k \in \mathbb{N}} \|x^{k+1} - x^k\|^2 < +\infty$;
- (ii) $(x^k)_{k \in \mathbb{N}}$ converges weakly to a point in $\text{Fix}(T)$.

3 An inertial primal-dual fixed point algorithm

3.1 Derivation of the algorithm

In the paper [5], Nesterov proposed a modification of the heavy ball method in order to improve the convergence rate on smooth convex functions. The idea of Nesterov was to use the extrapolated point y^k for evaluating the gradient. Moreover, in order to prove optimal convergence rates of the scheme, the extrapolation parameter α_k must satisfy some special conditions. The scheme is given by:

$$\begin{cases} l^k = x^k + \alpha_k(x^k - x^{k-1}), \\ x^{k+1} = l^k - \bar{\lambda}_k \nabla f(l^k), \end{cases} \quad (3.1)$$

where $\bar{\lambda}_k = 1/L$, there are several choices to define an optimal sequence α_k [5-8].

For Problem (1.4), When \bar{G} is differentiable on \mathcal{X} and its gradient $\nabla \bar{G}$ is β -Lipschitz continuous, for some $\beta \in [0, +\infty[$. Laurent Condat [1] give the following method:

Choose $x^0 \in \mathcal{X}$, $y^0 \in \mathcal{Y}$, relaxation parameters $(\rho_k)_{k \in \mathbb{N}}$, and proximal parameters $\sigma > 0$, $\tau > 0$. The iterate, for every $k \geq 0$

$$\begin{cases} \tilde{y}^{k+1} = \text{prox}_{\sigma F^*}(y^k + \sigma Kx^k), \\ \tilde{x}^{k+1} = \text{prox}_{\tau H}(x^k - \tau \nabla \bar{G}(x^k) - \tau K^*(2\tilde{y}^{k+1} - y^k)), \\ (x^{k+1}, y^{k+1}) = \rho_k(\tilde{x}^{k+1}, \tilde{y}^{k+1}) + (1 - \rho_k)(x^k, y^k). \end{cases} \quad (3.2)$$

Based on the idea of Laurent Condat[1] and Nesterov[5], we introduce the following new algorithm for solving Problem (1.4).

Algorithm 1 An inertial primal-dual fixed point algorithm(IPDFP).

Initialization: Choose $x^0, x^1, y^0, y^1 \in \mathcal{X}$, $v^0, v^1 \in \mathcal{Y}$, relaxation parameters $(\rho_k)_{k \in \mathbb{N}}$, extrapolation parameter α_k and proximal parameters $\sigma > 0, \gamma > 0$, $\tau > 0$.

Iterations ($k \geq 0$): Update x^k, y^k, v^k as follows

$$\left\{ \begin{array}{l} \xi^k = x^k + \alpha_k(x^k - x^{k-1}), \\ \eta^k = y^k + \alpha_k(y^k - y^{k-1}), \\ \nu^k = v^k + \alpha_k(v^k - v^{k-1}), \\ \tilde{x}^{k+1} = \text{prox}_{\sigma H}(\xi^k - \sigma \eta^k - \sigma K^* \nu^k), \\ \tilde{y}^{k+1} = \text{prox}_{\gamma \bar{G}^*}(\eta^k + \gamma \xi^k), \\ \tilde{v}^{k+1} = \text{prox}_{\tau F^*}(\nu^k + \tau K(2\tilde{x}^{k+1} - \eta^k)), \\ (x^{k+1}, y^{k+1}, v^{k+1}) = \rho_k(\tilde{x}^{k+1}, \tilde{y}^{k+1}, \tilde{v}^{k+1}) + (1 - \rho_k)(x^k, y^k, v^k). \end{array} \right.$$

End for

Theorem 3.1. Let $\sigma > 0$, $\gamma > 0$, $\tau > 0$, $(\alpha_k)_{k \in \mathbb{N}}$ and the sequences $(\rho_k)_{k \in \mathbb{N}}$, be the parameters of Algorithms 1. Let the following conditions hold:

- (i) $\sigma\gamma + \sigma\tau\|K\|^2 < 1$,
 - (ii) $(\alpha_k)_{k \in \mathbb{N}}$ is nondecreasing with $\alpha_1 = 0$ and $0 \leq \alpha_k \leq \alpha < 1$ for every $k \geq 1$ and $\rho, \theta, \hat{\delta} > 0$ are such that $\hat{\delta} > \frac{\alpha^2(1+\alpha)+\alpha\theta}{1-\alpha^2}$ and $0 < \rho \leq \rho_k < \frac{\hat{\delta}-\alpha[\alpha(1+\alpha)+\alpha\hat{\delta}+\theta]}{\hat{\delta}[1+\alpha(1+\alpha)+\alpha\hat{\delta}+\theta]} \forall k \geq 1$.
- Let the sequences (x^k, y^k, v^k) be generated by Algorithms 1. Then the sequence $\{x_k\}$ converges to a solution of Problem (1.4).

In the following, we would like to extend the IPDFP to solve the optimization problem (1.3).

Algorithm 2 A splitting inertial primal-dual fixed point algorithm(SIPDFP).

Initialization: Choose $x^0, x^1, y^0, y^1 \in \mathcal{X}$, $v_1^0, v_1^1 \in \mathcal{Y}_1, \dots, v_m^0, v_m^1 \in \mathcal{Y}_m$, relaxation parameters $(\rho_k)_{k \in \mathbb{N}}$, extrapolation parameter α_k and proximal parameters $\sigma > 0, \gamma > 0, \tau > 0$.

Iterations: for every $k \geq 0$

$$\left\{ \begin{array}{l} \xi^k = x^k + \alpha_k(x^k - x^{k-1}), \\ \tilde{x}^{k+1} = \xi^k - \sigma\eta^k - \frac{\sigma}{m} \sum_{i=1}^m K_i^* \nu_i^k, \\ x^{k+1} = \rho_k \tilde{x}^{k+1} + (1 - \rho_k)x^k, \\ \eta^k = y^k + \alpha_k(y^k - y^{k-1}), \\ \tilde{y}^{k+1} = \text{prox}_{\gamma G^*}(\eta^k + \gamma\xi^k), \\ y^{k+1} = \rho_k \tilde{y}^{k+1} + (1 - \rho_k)y^k, \\ \nu_i^k = v_i^k + \alpha_k(v_i^k - v_i^{k-1}), i = 1, \dots, m, \\ \tilde{v}_i^{k+1} = \text{prox}_{\tau F_i^*}(\nu_i^k + \tau K_i(2\tilde{x}^{k+1} - \eta^k)), i = 1, \dots, m, \\ v_i^{k+1} = \rho_k \tilde{v}_i^{k+1} + (1 - \rho_k)v_i^k, i = 1, \dots, m. \end{array} \right.$$

End for

Theorem 3.2. Let $\sigma > 0, \gamma > 0, \tau > 0$, $(\alpha_k)_{k \in \mathbb{N}}$ and the sequences $(\rho_k)_{k \in \mathbb{N}}$, be the parameters of Algorithms 2. Let the following conditions hold:

- (i) $\sigma\gamma + \sigma\tau \sum_{i=1}^m \|K_i\|^2 < 1$,
- (ii) $(\alpha_k)_{k \in \mathbb{N}}$ is nondecreasing with $\alpha_1 = 0$ and $0 \leq \alpha_k \leq \alpha < 1$ for every $k \geq 1$ and $\rho, \theta, \hat{\delta} > 0$ are such that $\hat{\delta} > \frac{\alpha^2(1+\alpha)+\alpha\theta}{1-\alpha^2}$ and $0 < \rho \leq \rho_k < \frac{\hat{\delta}-\alpha[\alpha(1+\alpha)+\alpha\hat{\delta}+\theta]}{\hat{\delta}[1+\alpha(1+\alpha)+\alpha\hat{\delta}+\theta]} \forall k \geq 1$.

Let the sequences (x^k, y^k, v^k) be generated by Algorithms 2. Then the sequence $\{x_k\}$ converges to a solution of Problem (1.3).

3.2 Proofs of convergence

Proof of Theorem 3.1 for Algorithm 1. By the idea of Laurent Condat[1], we know that Algorithm 1 has the structure of a forward-backward iteration, when expressed in terms of nonexpansive operators on $\mathcal{Z} := \mathcal{X} \times \mathcal{X} \times \mathcal{Y}$, equipped with a particular inner product.

Let the inner product $\langle \cdot, \cdot \rangle_I$ in \mathcal{Z} be defined as

$$\langle z, z' \rangle := \langle x, x' \rangle + \langle y, y' \rangle + \langle v, v' \rangle, \quad \forall z = (x, y, v), \quad z' = (x', y', v') \in \mathcal{Z}.$$

By endowing \mathcal{Z} with this inner product, we obtain the Euclidean space denoted by \mathcal{Z}_I . Let us define the bounded linear operator on \mathcal{Z} ,

$$P := \begin{pmatrix} x \\ y \\ v \end{pmatrix} \mapsto \begin{pmatrix} \frac{1}{\sigma} & -I & -K^* \\ -I & \frac{1}{\gamma} & 0 \\ -K & 0 & \frac{1}{\tau} \end{pmatrix} \begin{pmatrix} x \\ y \\ v \end{pmatrix}. \quad (3.3)$$

From the condition (i), we can easily check that P is positive definite. Hence, we can define another inner product $\langle \cdot, \cdot \rangle_P$ and norm $\| \cdot \|_P = \langle \cdot, \cdot \rangle_P^{\frac{1}{2}}$ in \mathcal{Z} as

$$\langle z, z' \rangle_P = \langle z, Pz' \rangle_I. \quad (3.4)$$

We denote by \mathcal{Z}_P the corresponding Euclidean space.

For every $k \in \mathbb{N}$, the following inclusion is satisfied by $\tilde{z}^{k+1} := (\tilde{x}^{k+1}, \tilde{y}^{k+1}, \tilde{v}^{k+1})$ computed by Algorithms 1:

$$0 \in \begin{pmatrix} \partial H & I & K^* \\ -I & \partial G^* & 0 \\ -K & 0 & \partial F^* \end{pmatrix} \begin{pmatrix} \tilde{x}^{k+1} \\ \tilde{y}^{k+1} \\ \tilde{v}^{k+1} \end{pmatrix} + \begin{pmatrix} \frac{1}{\sigma} & -I & -K^* \\ I & \frac{1}{\gamma} & 0 \\ -K & 0 & \frac{1}{\tau} \end{pmatrix} \begin{pmatrix} \tilde{x}^{k+1} - \xi^k \\ \tilde{y}^{k+1} - \eta^k \\ \tilde{v}^{k+1} - \nu^k \end{pmatrix}$$

By set

$$\varpi^k = (\xi^k, \eta^k, \nu^k), A := \begin{pmatrix} \partial H & I & K^* \\ -I & \partial G^* & 0 \\ -K & 0 & \partial F^* \end{pmatrix},$$

it also can be written as follows:

$$\begin{cases} \varpi^k = z^k + \alpha_k(z^k - z^{k-1}), \\ \tilde{z}^{k+1} := (I + P^{-1} \circ A)^{-1}(\varpi^k). \end{cases} \quad (3.5)$$

Considering the relaxation step, we obtain

$$\begin{cases} \varpi^k = z^k + \alpha_k(z^k - z^{k-1}), \\ \tilde{z}^{k+1} := (I + P^{-1} \circ A)^{-1}(\varpi^k), \\ z^{k+1} := \rho_k(I + P^{-1} \circ A)^{-1}(\varpi^k) + (1 - \rho_k)\varpi^k. \end{cases} \quad (3.6)$$

Set $M = P^{-1} \circ A$, then

The operator $(x, y, v) \mapsto \partial H \times \partial G^* \times \partial F^*$ is maximally monotone in \mathcal{Z}_I by Propositions 23.16 of [3]. Moreover, the skew operator:

$$\begin{pmatrix} x \\ y \\ v \end{pmatrix} \mapsto \begin{pmatrix} 0 & I & K^* \\ -I & 0 & 0 \\ -K & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ v \end{pmatrix}$$

is maximally monotone in \mathcal{Z}_I [3, Example 20.30] and has full domain. Hence, A is maximally monotone [3, Corollary 24.4(i)]. Thus, M is monotone in \mathcal{Z}_P and, from the injectivity of P , M is maximally monotone in \mathcal{Z}_P . Set $T = (I + M)^{-1}$, then by [3, Corollary 23.8], we know that $T \in \mathcal{A}(\mathcal{Z}_P, \frac{1}{2})$. In particular, it is non-expansive. Since P^{-1} and L are bounded and the norms $\|\cdot\|_I$ and $\|\cdot\|_P$ are equivalent, so from conditions (i)-(ii) and Lemma 2.2 we have that the iterative scheme defined by (3.6) satisfies the following statements:

- (i) $\sum_{k \in \mathbb{N}} \|z^{k+1} - z^k\|_P^2 < +\infty$;
- (ii) $(z^k)_{k \in \mathbb{N}}$ converges to a point in $\text{Fix}(T)$.

Then the sequence $\{x^k\}$ converges to a solution of Problem (1.4).

Elaborating on Theorem 3.1, we are now ready to establish the Theorem 3.2.

By the notation in Section 1, we know that, for any $\mathbf{y} = (y_1, \dots, y_m) \in \mathcal{Y}_1 \times \dots \times \mathcal{Y}_m$, $F^*(y_1, \dots, y_m) = (F_1^*y_1, \dots, F_m^*y_m)$, $K^*(y_1, \dots, y_m) = (K_1^*y_1, \dots, K_m^*y_m)$, $\text{prox}_{\tau F^*} = (\text{prox}_{\tau F_1^*}(y_1), \dots, \text{prox}_{\tau F_m^*}(y_m))$, $\|K\|^2 = \|\sum_{i=1}^m K_i^* K_i\|$. When $H(x) = \delta_C(x)$, where C is the space of vectors $x \in \mathcal{X}^m$, we know that for any $x \in \mathcal{X}^m$, $\text{proj}_C(x) = (\bar{x}, \dots, \bar{x})$ where \bar{x} is the average of vector x , i.e., $\bar{x} = m^{-1} \sum_{i=1}^m x_i$. Consequently, the components of \tilde{x}^{k+1} in Algorithm 1 are equal and coincide with $\xi^k - \sigma\eta^k - \frac{\sigma}{m} \sum_{i=1}^m K_i^* \nu_i^k$. Therefore, we can obtain Algorithm 2 by Algorithm 1, and we can obtain the convergence of Theorem 3.2 directly by Theorem 3.1.

3.3 Connections to other algorithms

We will further establish the connections to other existing methods.

Primal-dual algorithms

If the term \bar{G} is absent of the Problem (1.4), and $\alpha_k \equiv 0$, the Algorithms 1 boils down

to the primal-dual algorithms of Chambolle and Pock [12], which have been proposed in other forms in [13, 14].

Forward-backward splitting

If the term $F \circ K = 0$, \bar{G} is differentiable and its gradient $\nabla \bar{G}$ is Lipschitz continuous, $\alpha_k = 0$ in Algorithms 1. We obtain exactly the popular forward-backward splitting algorithm for minimizing the sum of a smooth and a non-smooth convex function. See [15, 16].

4 Preconditioning

4.1 Convergence of the Preconditioned algorithm

In the context of saddle point problems, Pock and Chambolle [17] proposed a preconditioning of the form

$$B := \begin{pmatrix} \tilde{T}^{-1} & -K^* \\ -K & \Sigma^{-1} \end{pmatrix}$$

where \tilde{T} and Σ are selfadjoint, positive definite maps. A condition for the positive definiteness of P follows from the following Lemma.

Lemma 4.1. ([14]). *Let A_1, A_2 be symmetric positive definite maps and M a bounded operator. If $\|A_2^{-\frac{1}{2}} M A_1^{-\frac{1}{2}}\| < 1$, then*

$$A := \begin{pmatrix} A_1 & M^* \\ M & A_2 \end{pmatrix}$$

is positive definite.

Based on the idea of Pock and Chambolle, we present a preconditioning of the form

$$\bar{P} := \begin{pmatrix} \Sigma^{-1} & -I & -K^* \\ -I & \Upsilon^{-1} & 0 \\ -K & 0 & \tilde{T}^{-1} \end{pmatrix}, \quad (4.1)$$

where Σ , Υ and \tilde{T} are selfadjoint, positive definite maps. A condition for the positive definiteness of \bar{P} follows from the following Lemma.

Lemma 4.2. *Let Σ , Υ and \tilde{T} be symmetric positive definite maps and \bar{P} a bounded operator. If $\|\Sigma^{\frac{1}{2}}\Upsilon^{\frac{1}{2}}\|^2 + \|\Sigma^{\frac{1}{2}}K^*\tilde{T}^{\frac{1}{2}}\|^2 < 1$, then the matrix \bar{P} defined in (4.1) is symmetric and positive definite .*

Proof. Due to the structure of \bar{P} , we have that

$$\left\langle \begin{pmatrix} x \\ y \\ v \end{pmatrix}, \bar{P} \begin{pmatrix} x \\ y \\ v \end{pmatrix} \right\rangle = \langle x, \Sigma^{-1}x \rangle + \langle y, \Upsilon^{-1}y \rangle + \langle v, \tilde{T}^{-1}v \rangle - 2\langle y + K^*v, x \rangle,$$

set

$$D = (I, K^*), u = (y, v)^T, M := \begin{pmatrix} \Upsilon & 0 \\ 0 & \tilde{T} \end{pmatrix}$$

then estimate the middle term from below by Cauchy- Schwarz and Youngs inequality and get for every $\varepsilon > 0$ that

$$\begin{aligned} -2\langle y + K^*v, x \rangle &= -2\langle Du, x \rangle = -2\langle \Sigma^{\frac{1}{2}}DM^{\frac{1}{2}}M^{-\frac{1}{2}}u, \Sigma^{-\frac{1}{2}}x \rangle \\ &\geq -2\|\Sigma^{\frac{1}{2}}DM^{\frac{1}{2}}M^{-\frac{1}{2}}u\|\|\Sigma^{-\frac{1}{2}}x\| \\ &\geq -(\varepsilon\|\Sigma^{\frac{1}{2}}DM^{\frac{1}{2}}\|^2\|u\|_{M^{-\frac{1}{2}}}^2 + \frac{1}{\varepsilon}\|x\|_{\Sigma^{-\frac{1}{2}}}^2). \end{aligned}$$

Since $\|\Sigma^{\frac{1}{2}}DM^{\frac{1}{2}}\|^2 = \|\Sigma^{\frac{1}{2}}\Upsilon^{\frac{1}{2}}\|^2 + \|\Sigma^{\frac{1}{2}}K^*\tilde{T}^{\frac{1}{2}}\|^2 < 1$, so we have

$$\left\langle \begin{pmatrix} x \\ y \\ v \end{pmatrix}, \bar{P} \begin{pmatrix} x \\ y \\ v \end{pmatrix} \right\rangle \geq (1 - \varepsilon\|\Sigma^{\frac{1}{2}}DM^{\frac{1}{2}}\|^2)\|u\|_{M^{-\frac{1}{2}}}^2 + (1 - \frac{1}{\varepsilon})\|x\|_{\Sigma^{-\frac{1}{2}}}^2 > 0.$$

□

Now, we study preconditioning techniques for the inertial primal-dual fixed point algorithm(IPDFP), then we obtain the following algorithm.

Algorithm 3 An inertial primal-dual fixed point algorithm with preconditioning (IPDFP²).

Initialization: Choose $x^0, x^1, y^0, y^1 \in \mathcal{X}$, $v^0, v^1 \in \mathcal{Y}$, relaxation parameters $(\rho_k)_{k \in \mathbb{N}}$, extrapolation parameter α_k and positive definite maps Σ , Υ and \tilde{T} .
Iterations ($k \geq 0$): Update x^k, y^k, v^k as follows

$$\left\{ \begin{array}{l} \xi^k = x^k + \alpha_k(x^k - x^{k-1}), \\ \eta^k = y^k + \alpha_k(y^k - y^{k-1}), \\ \nu^k = v^k + \alpha_k(v^k - v^{k-1}), \\ \tilde{x}^{k+1} = \text{prox}_{\Sigma H}(\xi^k - \Sigma \eta^k - \Sigma K^* \nu^k), \\ \tilde{y}^{k+1} = \text{prox}_{\Upsilon \bar{G}^*}(\eta^k + \Upsilon \xi^k), \\ \tilde{v}^{k+1} = \text{prox}_{\tilde{T} F^*}(\nu^k + \tilde{T} K(2\tilde{x}^{k+1} - \eta^k)), \\ (x^{k+1}, y^{k+1}, v^{k+1}) = \rho_k(\tilde{x}^{k+1}, \tilde{y}^{k+1}, \tilde{v}^{k+1}) + (1 - \rho_k)(x^k, y^k, v^k). \end{array} \right.$$

End for

It turns out that the resulting method converges under appropriate conditions.

Theorem 4.1. *In the setting of Theorem 3.1 let the following conditions holds :*

- (i) $\|\Sigma^{\frac{1}{2}} \Upsilon^{\frac{1}{2}}\|^2 + \|\Sigma^{\frac{1}{2}} K^* \tilde{T}^{\frac{1}{2}}\|^2 < 1$;
- (ii) $(\alpha_k)_{k \in \mathbb{N}}$ is nondecreasing with $\alpha_1 = 0$ and $0 \leq \alpha_k \leq \alpha < 1$ for every $k \geq 1$ and $\rho, \theta, \hat{\delta} > 0$ are such that $\hat{\delta} > \frac{\alpha^2(1+\alpha)+\alpha\theta}{1-\alpha^2}$ and $0 < \rho \leq \rho_k < \frac{\hat{\delta}-\alpha[\alpha(1+\alpha)+\alpha\hat{\delta}+\theta]}{\hat{\delta}[1+\alpha(1+\alpha)+\alpha\hat{\delta}+\theta]} \forall k \geq 1$.
Then the sequence $\{x^k\}$ generated by the Algorithm 3 converges to a solution of Problem (1.4).

Proof. As shown in Lemma 4.2, the condition $\|\Sigma^{\frac{1}{2}} \Upsilon^{\frac{1}{2}}\|^2 + \|\Sigma^{\frac{1}{2}} K^* \tilde{T}^{\frac{1}{2}}\|^2 < 1$ ensure that the matrix \bar{P} defined in (4.1) is symmetric and positive definite. Therefore, with the same proof of Theorem 3.1, we can obtain Theorem 4.1. □

For selfadjoint, positive definite maps Σ , Υ , \tilde{T} , we consider the following algorithm which we shall refer to as a preconditioned splitting inertial primal-dual fixed point algorithm(PSIPDFP).

Algorithm 4 Preconditioned splitting inertial primal-dual fixed point algorithm(PSIPDFP).

Initialization: Choose $x^0, x^1, y^0, y^1 \in \mathcal{X}$, $v_1^0, v_1^1 \in \mathcal{Y}_1, \dots, v_m^0, v_m^1 \in \mathcal{Y}_m$, relaxation parameters $(\rho_k)_{k \in \mathbb{N}}$, extrapolation parameter α_k and positive definite maps Σ , Υ and \tilde{T} .

Iterations: for every $k \geq 0$

$$\left\{ \begin{array}{l} \xi^k = x^k + \alpha_k(x^k - x^{k-1}), \\ \tilde{x}^{k+1} = \xi^k - \Sigma\eta^k - \frac{1}{m}\Sigma \sum_{i=1}^m K_i^* \nu_i^k, \\ x^{k+1} = \rho_k \tilde{x}^{k+1} + (1 - \rho_k)x^k, \\ \eta^k = y^k + \alpha_k(y^k - y^{k-1}), \\ \tilde{y}^{k+1} = \text{prox}_{\Upsilon G^*}(\eta^k + \Upsilon \xi^k), \\ y^{k+1} = \rho_k \tilde{y}^{k+1} + (1 - \rho_k)y^k, \\ \nu_i^k = v_i^k + \alpha_k(v_i^k - v_i^{k-1}), i = 1, \dots, m, \\ \tilde{v}_i^{k+1} = \text{prox}_{\tilde{T} F_i^*}(\nu_i^k + \tilde{T} K_i(2\tilde{x}^{k+1} - \eta^k)), i = 1, \dots, m, \\ v_i^{k+1} = \rho_k \tilde{v}_i^{k+1} + (1 - \rho_k)v_i^k, i = 1, \dots, m. \end{array} \right.$$

End for

Theorem 4.2. *In the setting of Theorem 3.2 let the following conditions holds :*

- (i) $\|\Sigma^{\frac{1}{2}}\Upsilon^{\frac{1}{2}}\|^2 + \sum_{i=1}^m \|\Sigma^{\frac{1}{2}}K_i^*\tilde{T}^{\frac{1}{2}}\|^2 < 1$;
- (ii) $(\alpha_k)_{k \in \mathbb{N}}$ is nondecreasing with $\alpha_1 = 0$ and $0 \leq \alpha_k \leq \alpha < 1$ for every $k \geq 1$ and $\rho, \theta, \hat{\delta} > 0$ are such that $\hat{\delta} > \frac{\alpha^2(1+\alpha)+\alpha\theta}{1-\alpha^2}$ and $0 < \rho \leq \rho_k < \frac{\hat{\delta}-\alpha[\alpha(1+\alpha)+\alpha\hat{\delta}+\theta]}{\hat{\delta}[1+\alpha(1+\alpha)+\alpha\hat{\delta}+\theta]} \forall k \geq 1$.

Then the sequence $\{x^k\}$ generated by the Algorithm 4 converges to a solution of Problem (1.3).

Proof. Set

$$\bar{K}^* := \begin{pmatrix} K_1^* \\ \vdots \\ K_m^* \end{pmatrix}, \bar{\mathbf{P}} := \begin{pmatrix} \Sigma^{-1} & -I & -\bar{K}^* \\ -I & \Upsilon^{-1} & 0 \\ -\bar{K} & 0 & \tilde{T}^{-1} \end{pmatrix}.$$

Then from the condition $\|\Sigma^{\frac{1}{2}}\Upsilon^{\frac{1}{2}}\|^2 + \sum_{i=1}^m \|\Sigma^{\frac{1}{2}}K_i^*\tilde{T}^{\frac{1}{2}}\|^2 < 1$ and Lemma 4.2, we can know that $\bar{\mathbf{P}}$ is symmetric and positive definite. Hence, the convergence of the

Algorithm 4 to an optimal solution of (1.3) follows from the weak convergence of the Algorithm 2. □

4.2 Diagonal Preconditioning

In this section, we show how we can choose pointwise step sizes for both the primal and the dual variables that will ensure the convergence of the algorithm. The next result is an adaption of the preconditioner proposed in [17].

Lemma 4.3. *Let $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$, $\Upsilon = \text{diag}(\gamma_1, \dots, \gamma_n)$ and $\tilde{T} = \text{diag}(\tau_1, \dots, \tau_m)$, then we can know that $M = \text{diag}(\gamma_1, \dots, \gamma_n, \tau_1, \dots, \tau_m)$. In particular, we set $M = \text{diag}(\gamma_1, \dots, \gamma_n, \tau_1, \dots, \tau_m) = \text{diag}(\varphi_1, \dots, \varphi_{n+m})$ with*

$$\sigma_j = \frac{1}{\sum_{i=1}^{n+m} |D_{i,j}|^{2-s}}, \varphi_i = \frac{1}{\sum_{j=1}^n |D_{i,j}|^s}, \quad (4.2)$$

then for any $s \in [0, 2]$

$$\|\Sigma^{\frac{1}{2}} \Upsilon^{\frac{1}{2}}\|^2 + \|\Sigma^{\frac{1}{2}} K^* \tilde{T}^{\frac{1}{2}}\|^2 = \|\Sigma^{\frac{1}{2}} D M^{\frac{1}{2}}\|^2 \leq 1. \quad (4.3)$$

Proof. In order to prove the inequality, we need to find an upper bound on $\|\Sigma^{\frac{1}{2}} D M^{\frac{1}{2}}\|^2$. From the proof of [17, Lemma 2] we can obtain the results directly. □

5 Numerical experiments

We consider the problem of l_1 -regularized logistic regression. Denoting by m the number of observations and by q the number of features, the optimization problem writes

$$\inf_{x \in \mathbb{R}^q} \frac{1}{m} \sum_{i=1}^m \log(1 + e^{-y_i a_i^T x}) + \tau \|x\|_1, \quad (8.1)$$

where the $(y_i)_{i=1}^m$ are in $\{-1, +1\}$, the $(a_i)_{i=1}^m$ are in \mathbb{R}^q , and $\tau > 0$ is a scalar. Let $(\mathcal{W})_{n=1}^N$ indicate a partition of $\{1, \dots, m\}$. The optimization problem then writes

$$\inf_{x \in \mathbb{R}^q} \sum_{n=1}^N \sum_{i \in \mathcal{W}_n} \frac{1}{m} \log(1 + e^{-y_i a_i^T x}) + \tau \|x\|_1, \quad (8.2)$$

or, splitting the problem between the batches

$$\inf_{x \in \mathbb{R}^{Nq}} \sum_{n=1}^N \left(\sum_{i \in \mathcal{W}_n} \frac{1}{m} \log(1 + e^{-y_i a_i^T x_n}) + \frac{\tau}{N} \|x_n\|_1 \right) + \iota_{\mathcal{C}}(x), \quad (8.3)$$

where $x = (x_1, \dots, x_N)$ is in \mathbb{R}^{Nq} . It is easy to see that Problems (8.1), (8.2) and (8.3) are equivalent and Problem (8.3) is in the form of (6.2).

6 Conclusion

In this paper, we introduced a new framework for stochastic coordinate descent and used on a algorithm called ADMMS⁺. As a byproduct, we obtained a stochastic approximation algorithm with dynamic stepsize which can be used to handle distinct data blocks sequentially. We also obtained an asynchronous distributed algorithm with dynamic stepsize which enables the processing of distinct blocks on different machines.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (11131006, 41390450, 91330204, 11401293), the National Basic Research Program of China (2013CB 329404), the Natural Science Foundations of Jiangxi Province (CA201107114, 20114BAB 201004).

References

- [1] L. Condat, A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms, *Journal of Optimization Theory and Applications*, vol. 158, no. 2, pp. 460C479, 2013.
- [2] F. Alvarez, Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in Hilbert space, *SIAM Journal on Optimization* 14(3), 773-782, 2004

- [3] Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York (2011)
- [4] BoT, R. I., Csetnek, E. R. and Hendrich, C.: Inertial Douglas-Rachford splitting for monotone inclusion problems. arXiv:1403.3330v2 [math.OC] 28 Mar 201
- [5] Nesterov, Yu.: A method for solving the convex programming problem with convergence rate $O(1/k^2)$. Dokl. Akad. Nauk SSSR 269(3), 543-547 (1983)
- [6] Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. 2(1), 183-202 (2009)
- [7] Nesterov, Y.: Introductory lectures on convex optimization: a basic course. In: Applied Optimization, vol. 87. Kluwer Academic Publishers, Boston, MA (2004)
- [8] Tseng, P.: On accelerated proximal gradient methods for convexconcave optimization. Technical report (2008)
- [9] Bauschke HH, Combettes PL. Convex Analysis and Motonone Operator Theory in Hilbert Spaces[M]. Springer, London, 2011.
- [10] F. Alvarez, On the minimizing property of a second order dissipative system in Hilbert spaces, SIAM Journal on Control and Optimization 38(4), 1102-1119, 2000
- [11] I. Daubechies, M. Defrise, and C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Communications on pure and applied mathematics, vol. 57, no. 11, pp. 1413-1457, 2004.
- [12] Chambolle, A., Pock, T.: A first-order primalCdual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. 40(1), 120-145 (2011)
- [13] Esser, E., Zhang, X., Chan, T.: A general framework for a class of first order primalCdual algorithms for convex optimization in imaging science. SIAM J. Imaging Sci. 3(4), 1015C1046 (2010)
- [14] Zhang, X., Burger, M., Osher, S.: A unified primal-dual algorithm framework based on Bregman iteration. J. Sci. Comput. 46(1), 20C46 (2010)

- [15] Combettes, P.L., Wajs, V.: Signal recovery by proximal forwardbackward splitting. *SIAM Multiscale Model. Simul.* 4(4), 1168- 1200 (2005)
- [16] Daubechies, I., Defrise, M., De Mol, C.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.* 57, 1413-1457 (2004)
- [17] Pock, T., Chambolle, A.: Diagonal preconditioning for first order primal-dual algorithms. In: *Proceedings of the International Conference of Computer Vision (ICCV 2011)*, pp. 1762-1769 (2011)
- [18] Lorenz Dirk A., Pock T. An inertial forward-backward algorithm for monotone inclusions. *J. Math Imaging Vis* DOI 10.1007/s10851-014-0523-2.
- [19] Chambolle, A., Lions, P.L.: Image recovery via total variational minimization and related problems. *Numer. Math.* 76, 167-188 (1997)
- [20] Chan, T., Marquina, A., Mulet, P.: High-order total variation-based image restoration. *SIAM J. Sci. Comput.* 22, 503-516 (2000)
- [21] Lysaker, M., Tai, X.: Iterative image restoration combining total variation minimization and a second-order functional. *Int. J. Comput. Vis.* 66, 5-18 (2006)
- [22] Scherzer, O.: Denoising with higher-order derivatives of bounded variation and an application to parameter estimation. *Computing* 60, 5-27 (1998)
- [23] Stefan, W., Renaut, R.A., Gelb, A.: Improved total variation-type regularization using higher order edge detectors. *SIAM J. Imaging Sci.* 3, 232-251 (2010)
- [24] You, Y., Kaveh, M.: Fourth-order partial differential equations for noise removal. *IEEE Trans. Image Process.* 9, 1723-1730 (2000)